Dually flat structure with escort probability and its application to alpha-Voronoi diagrams‡

Atsumi Ohara[1], Hiroshi Matsuzoe[2] and Shun-ichi Amari[3]

[1]Department of Systems Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

E-mail: `ohara@sys.es.osaka-u.ac.jp`

[2]Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan

E-mail: `matsuzoe@nitech.ac.jp`

[3]Riken Brain Science Institute, Wako-shi Hirosawa 2-1, Saitama 351-0198 Japan

E-mail: `amari@brain.riken.jp`

Abstract. This paper studies geometrical structure of the manifold of escort probability distributions and shows its new applicability to information science. In order to realize escort probabilities we use a conformal transformation that flattens so-called alpha-geometry of the space of discrete probability distributions, which well characterizes nonadditive statistics on the space. As a result escort probabilities are proved to be flat coordinates of the usual probabilities for the derived dually flat structure. Finally, we demonstrate that escort probabilities with the new structure admits a simple algorithm to compute Voronoi diagrams and centroids with respect to alpha-divergences.

‡ Several results in this paper can be found in the conference paper [36] without complete proofs.

1. Introduction

Escort probability is naturally induced from researches of multifractals [1] and non-extensive statistical mechanics [2] to play an important but mysterious role. Testing its utility in the other scientific fields would greatly help our understanding about it. This motivates us to approach the escort probability by geometrically studying its role in information science.

The first purpose of this paper is to investigate the escort probability from viewpoints of information geometry [3, 4] and affine differential geometry [5]. The second is to show that escort probability with information geometric structure is useful to construction of Voronoi diagrams [6] on the space of probability distributions.

Recently, it is reported [7, 8] that α-*geometry*, which is an information geometric structure of constant curvature, has a close relation with Tsallis statistics [2]. The remarkable feature of the α-geometry consists of the Fisher metric together with a one-parameter family of dual affine connections, called the α-connections.

We prove that the manifold of escort probability distributions is dually flat by considering conformal transformations that flatten the α-geometry on the manifold of usual probability distributions. On the resultant manifold, escort probabilities consist of an affine coordinate system. See also [9] for another type of flattening a curved dual manifold by a conformal transformation.

The result gives us a clear geometrical interpretation of the escort probability, and simultaneously, produces its new obscure links to conformality and projectivity. Due to these two geometrical concepts, however, the obtained dually flat structure inherits several properties of the α-geometry.

The dually flatness proves crucial to construction of Voronoi diagrams for α-divergences, which we shall call α-*Voronoi diagrams*. The Voronoi diagrams on the space of probability distributions with the Kullback-Leibler [10, 11], or Bregman divergences [12] have been recognized as important tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on. See also, e.g., [13, 14, 15] for related problems.

The largest advantage to take account of α-divergences is their *invariance under transformations by sufficient statistics* [16] (See also [4] in a different viewpoint), which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the α-geometry enables us to invoke the standard algorithm [29, 6] using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates.

Section 2 is devoted to preliminaries for α-geometry in the light of affine differential geometry. In section 3, as a main result, we consider conformal transformations and discuss properties of the obtained dually flat structure. Dual pairs of potential functions and affine coordinate systems on the manifold are explicitly identified, and the associated canonical divergence is shown to be conformal to the α-divergence. Section 4 describes an application of such a flattened geometric structure to α-Voronoi diagrams on the

probability simplex. The properties and a construction algorithm are discussed. Further, a formula for α-centroid is touched upon.

In the sequel, we fix the relations of two parameters q and α as $q = (1 - \alpha)/2$, and restrict $q > 0$.

2. Preliminaries

We briefly introduce α-geometry via affine differential geometry. See for details [7, 8]. Let $\mathcal{S}^n$ denote the n-dimensional probability simplex, i.e.,

$$\mathcal{S}^n := \left\{ \boldsymbol{p} = (p_i) \,\middle|\, p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right\}, \tag{1}$$

and $p_i, i = 1, \cdots, n + 1$ denote probabilities of $n + 1$ states. We introduce the α-geometric structure on $\mathcal{S}^n$. Let $\{\partial_i\}, i = 1, \cdots, n$ be natural basis tangent vector fields on $\mathcal{S}^n$ defined by

$$\partial_i := \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}, \quad i = 1, \cdots, n, \tag{2}$$

where $p_{n+1} = 1 - \sum_{i=1}^{n} p_i$. Now we define a Riemannian metric g on $\mathcal{S}^n$ called the *Fisher metric*:

$$g_{ij}(\boldsymbol{p}) := g(\partial_i, \partial_j) = \frac{1}{p_i}\delta_{ij} + \frac{1}{p_{n+1}} \tag{3}$$

$$= \sum_{k=1}^{n+1} p_k (\partial_i \log p_k)(\partial_j \log p_k), \quad i, j = 1, \cdots, n.$$

Further, define an torsion-free affine connection $\nabla^{(\alpha)}$ called the α-*connection*, which is represented in its coefficients by

$$\Gamma_{ij}^{(\alpha)k}(\boldsymbol{p}) = \frac{1+\alpha}{2}\left(-\frac{1}{p_k}\delta_{ij}^k + p_k g_{ij}\right), \quad i, j, k = 1, \cdots, n, \tag{4}$$

where δ_{ij}^k is equal to one if $i = j = k$ and zero otherwise. Then we have the α-covariant derivative $\nabla^{(\alpha)}$, which gives

$$\nabla_{\partial_i}^{(\alpha)} \partial_j = \sum_{k=1}^{n} \Gamma_{ij}^{(\alpha)k} \partial_k,$$

when it is applied to the vector fields ∂_i and ∂_j.

There are two specific features for the α-geometry on $\mathcal{S}^n$ defined in such a way. First, the triple $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ is a *statistical manifold* [17] (See appendix A for its definition), i.e., we can confirm that the following relation holds:

$$Xg(Y,Z) = g(\nabla_X^{(\alpha)}Y, Z) + g(Y, \nabla_X^{(-\alpha)}Z), \quad X, Y, Z \in \mathcal{X}(\mathcal{S}^n), \tag{5}$$

where $\mathcal{X}(\mathcal{S}^n)$ denotes the set of all tangent vector fields on $\mathcal{S}^n$. Two statistical manifolds $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ and $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$ are said *mutually dual*.

The other is that $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ is a manifold of constant curvature $\kappa = (1 - \alpha^2)/4$, i.e.,

$$R^{(\alpha)}(X,Y)Z = \kappa\{g(Y,Z)X - g(X,Z)Y\},$$

where $R^{(\alpha)}$ is the curvature tensor with respect to $\nabla^{(\alpha)}$. From this property the well-known nonadditive formula of the Tsallis entropy can be derived [7].

In [8] we have discussed the α-geometry on $\mathcal{S}^n$ from a viewpoint of *affine differential geometry* [5]. Consider the immersion f of $\mathcal{S}^n$ into $\mathbf{R}_+^{n+1}$ by

$$f : \boldsymbol{p} = (p_i) \mapsto \boldsymbol{x} = (x^i) = (L^{(\alpha)}(p_i)), \quad i = 1, \cdots, n+1, \tag{6}$$

where $(x^i), i = 1, \cdots, n+1$ is the canonical flat coordinate system of $\mathbf{R}^{n+1}$ and the function $L^{(\alpha)}$ is defined by

$$L^{(\alpha)}(t) := \frac{2}{1-\alpha} t^{(1-\alpha)/2} = \frac{1}{q} t^q.$$

Note that $f(\mathcal{S}^n)$ is a level hypersurface in the ambient space $\mathbf{R}_+^{n+1}$ represented by $\Psi(\boldsymbol{x}) = 2/(1+\alpha)$, where

$$\Psi(\boldsymbol{x}) := \frac{2}{\alpha+1} \sum_{i=1}^{n+1} \left(\frac{1-\alpha}{2} x^i \right)^{2/(1-\alpha)} = \frac{1}{1-q} \sum_{i=1}^{n+1} \left(qx^i \right)^{1/q}. \tag{7}$$

We choose a transversal vector ξ on the level hypersurface by

$$\xi := \sum_{i=1}^{n+1} \xi^i \frac{\partial}{\partial x^i}, \quad \xi^i = -q(1-q)x^i = -\kappa x^i. \tag{8}$$

Then we can confirm that the affine immersion (f, ξ) realizes the α-geometry on $\mathcal{S}^n$ [8]. Hence, it would be possible to develop theory of the α-geometry and Tsallis statistics with ideas of affine differential geometry [18].

Further, the *escort probability* [1] naturally appears in this setup. The escort probability $\boldsymbol{P} = (P_i)$ associated with $\boldsymbol{p} = (p_i)$ is the normalized version of $(p_i)^q$, and is defined by

$$P_i(\boldsymbol{p}) := \frac{(p_i)^q}{\sum_{j=1}^{n+1}(p_j)^q} = \frac{x^i}{Z_q}, \ i = 1, \cdots, n+1, \quad Z_q(\boldsymbol{p}) := \sum_{i=1}^{n+1} x^i(\boldsymbol{p}), \ \boldsymbol{x}(\boldsymbol{p}) \in f(\mathcal{S}^n). \tag{9}$$

Hence, the simplex $\mathcal{E}^n$ in the ambient space $\mathbf{R}_+^{n+1}$, i.e.,

$$\mathcal{E}^n := \left\{ \boldsymbol{x} = (x^i) \,\middle|\, \sum_{i=1}^{n+1} x^i = 1, \ x^i > 0 \right\}$$

represents the set of escort distributions $\boldsymbol{P}$.

Note that the element $\boldsymbol{x}^* = (x_i^*)$ in the dual space of $\mathbf{R}^{n+1}$ defined by

$$x_i^*(\boldsymbol{p}) := L^{(-\alpha)}(p_i) = \frac{1}{1-q}(p_i)^{1-q}, \quad i = 1, \cdots, n+1,$$

meets

$$x_i^*(\boldsymbol{p}) = \frac{\partial \Psi}{\partial x^i}(\boldsymbol{x}(\boldsymbol{p})).$$

Hence, it satisfies [8]

$$-\sum_{i=1}^{n+1} \xi^i(\boldsymbol{p})x_i^*(\boldsymbol{p}) = 1, \quad \sum_{i=1}^{n+1} x_i^*(\boldsymbol{p})X^i = 0, \tag{10}$$

for an arbitrary vector $X = \sum_{i=1}^{n+1} X^i \partial/\partial x^i$ at $\boldsymbol{x}(\boldsymbol{p})$ tangent to $f(\mathcal{S}^n)$. Thus, $-\boldsymbol{x}^*(\boldsymbol{p})$ can be interpreted as the *conormal map* [5].

3. A conformally and projectively flat geometric structure and escort probabilities

In this section we show a main result. For this purpose, we consider a conformal and projective transformation [19, 20, 21, 22] of the α-geometry to introduce a dually flat one. This flattening of the α-geometry conserves some of its properties. The escort probabilities (P_i) are found to represent one of mutually dual affine coordinate systems in the induced geometry. While the many functions or geometric quantities introduced in this section depend on the parameter α or q, we omit them for the brevity.

Let us define a function λ on $\mathcal{S}^n$ by

$$\lambda(\boldsymbol{p}) := \frac{1}{Z_q} = \frac{1}{\sum_{i=1}^{n+1} L^{(\alpha)}(p_i)},$$

which depends on α. Then, from (9) $\mathcal{E}^n$ is regarded as the image of $\mathcal{S}^n$ for another immersion $\tilde{f} := \lambda f$, i.e.,

$$\tilde{f} : \mathcal{S}^n \ni (p_i) \mapsto (P_i) \in \mathcal{E}^n, \quad i = 1, \cdots, n+1,$$

and $(P_1, \cdots, P_n)$ is interpreted as another coordinate system of $\mathcal{S}^n$. Note that the inverse mapping $\tilde{f}^{-1}$ is well-defined by

$$\tilde{f}^{-1} : (P_i) \mapsto (p_i) = \left(\frac{(P_i)^{1/q}}{\sum_{j=1}^{n+1} (P_j)^{1/q}} \right), \quad i = 1, \cdots, n+1.$$

It would be a natural way to introduce geometric structure on $\mathcal{E}^n$ (and hence on $\mathcal{S}^n$) via the affine immersion $(\tilde{f}, \tilde{\xi})$ by taking a suitable transversal vector $\tilde{\xi}$, similarly to the case of the α-geometry mentioned above. Since $\mathcal{E}^n$ is a part of a hyperplane in $\mathbf{R}^{n+1}$, the canonical affine connection of $\mathbf{R}^{n+1}$ induces a flat connection, denoted by $D^{(\mathrm{E})}$, on $\mathcal{E}^n$. However, for the same reason, we cannot define a Riemannian metric in this way§ because it vanishes on $\mathcal{E}^n$, regardless of any choice of the transversal vector $\tilde{\xi}$.

The idea we adopt here is to define a Riemannian metric by utilizing a property of $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ called -1-*conformal flatness*. Based on the results proved by Kurose [19, 20], we conclude that the manifold $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ is ± 1-conformally flat (See Appendix A for its definition) because it is a statistical manifold of constant curvature.

Actually, let ∇^* be the flat connection‖ on $\mathcal{S}^n$ defined with $D^{(\mathrm{E})}$ and the differential $\tilde{f}_*$ by

$$\tilde{f}_*(\nabla_X^* Y) = D^{(\mathrm{E})}_{\tilde{f}_* X} \tilde{f}_* Y, \quad X, Y \in \mathcal{X}(\mathcal{S}^n).$$

Then, we can prove that $\nabla^{(\alpha)}$ and ∇^* are *projectively equivalent* [5], i.e., it holds that

$$\nabla_X^* Y = \nabla_X^{(\alpha)} Y + d(\ln \lambda)(Y)X + d(\ln \lambda)(X)Y, \quad X, Y \in \mathcal{X}(\mathcal{S}^n). \tag{11}$$

Hence, if we define another Riemannian metric h on $\mathcal{S}^n$ by

$$h(X, Y) := \lambda g(X, Y), \quad X, Y \in \mathcal{X}(\mathcal{S}^n), \tag{12}$$

§ In affine differential geometry, a Riemannian metric is realized as the affine fundamental form of an affine immersion [5].
‖ For the sake of notational consistency with the existing literature, e.g., [3, 4], we first define ∇^*, and later ∇ as the dual of ∇^*.

then, $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ is -1-conformally equivalent to $(\mathcal{S}^n, h, \nabla^*)$ equipped with a flat connection ∇^*. Further, the manifold $(\mathcal{S}^n, h, \nabla^*)$ can be proved to be a statistical manifold (See Appendix B).

Using the conormal map $-\boldsymbol{x}^*(\boldsymbol{p})$, we can define the α-*divergence* as a *contrast function* (See Appendix A) inducing $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ as follows [20]:

$$D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = -\sum_{i=1}^{n+1} x_i^*(\boldsymbol{r})(x^i(\boldsymbol{p}) - x^i(\boldsymbol{r}))$$

$$= \langle -\boldsymbol{x}^*(\boldsymbol{r}), \boldsymbol{x}(\boldsymbol{p}) - \boldsymbol{x}(\boldsymbol{r}) \rangle = \frac{1}{\kappa} - \langle \boldsymbol{x}^*(\boldsymbol{r}), \boldsymbol{x}(\boldsymbol{p}) \rangle.$$

The statistical manifolds $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$ and $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ are dual in the sense of (5). Further, it is known [4] that there exists the unique affine flat connection ∇ on $\mathcal{S}^n$, dual with respect to (h, ∇^*). Then, according to [20], it is proved that $(\mathcal{S}^n, h, \nabla)$ is 1-conformally equivalent to $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$ and a contrast function ρ inducing (h, ∇, ∇^*) is given by scaling $D^{(-\alpha)}$ (See Appendix A) as follows:

$$\rho(\boldsymbol{p}, \boldsymbol{r}) = \lambda(\boldsymbol{r}) D^{(-\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = \frac{1}{Z_q(\boldsymbol{r})} D^{(-\alpha)}(\boldsymbol{p}, \boldsymbol{r})$$

$$= \frac{1}{Z_q(\boldsymbol{r})} \langle -\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{x}^*(\boldsymbol{p}) - \boldsymbol{x}^*(\boldsymbol{r}) \rangle = \langle -\boldsymbol{P}(\boldsymbol{r}), \boldsymbol{x}^*(\boldsymbol{p}) - \boldsymbol{x}^*(\boldsymbol{r}) \rangle. \quad (13)$$

We shall call ρ a *conformal divergence*.

Now, since $(\mathcal{S}^n, h, \nabla, \nabla^*)$ is a dually flat space, the standard result in [3, 4] suggests that there exist *mutually dual affine coordinate systems* $(\theta^1, \cdots, \theta^n)$ and $(\eta_1, \cdots, \eta_n)$, a *potential function* $\psi(\boldsymbol{\theta})$ and *its conjugate* $\psi^*(\boldsymbol{\eta})$ satisfying

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \psi^*}{\partial \eta_i}, \quad i = 1, \cdots, n. \quad (14)$$

They completely determine dually flat structure, i.e., the coefficients of h, ∇ and ∇^* are derived as the second and third derivatives of ψ or ψ^*, for example,

$$h_{ij} = h\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j}\right) = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}, \quad h^{ij} = h\left(\frac{\partial}{\partial \eta_i}, \frac{\partial}{\partial \eta_j}\right) = \frac{\partial^2 \psi^*}{\partial \eta_i \partial \eta_j},$$

$$\Gamma_{ijk} = h\left(\nabla_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^j}, \frac{\partial}{\partial \theta^k}\right) = 0, \quad \Gamma_{ijk}^* = h\left(\nabla^*_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^j}, \frac{\partial}{\partial \theta^k}\right) = \frac{\partial^3 \psi}{\partial \theta^i \partial \theta^j \partial \theta^k},$$

and so on. In order to identify ψ, ψ^*, θ^i and η_i explicitly *without integrating h_{ij} or h^{ij}*, we shall search for them by examining whether the conformal divergence ρ can be represented in the form of the *canonical divergence* [4], i.e.,

$$\rho(\boldsymbol{p}, \boldsymbol{r}) = \psi(\boldsymbol{\theta}(\boldsymbol{p})) + \psi^*(\boldsymbol{\eta}(\boldsymbol{r})) - \sum_{i=1}^n \theta^i(\boldsymbol{p}) \eta_i(\boldsymbol{r}). \quad (15)$$

with the constraints (14). If this is possible, we can directly prove from (A.4) and (A.5) that the obtained $\psi, \psi^*, (\theta^1, \cdots, \theta^n)$ and $(\eta_1, \cdots, \eta_n)$ are pairs of dual potential functions and affine coordinate systems associated with $(\mathcal{S}^n, h, \nabla, \nabla^*)$.

Before showing the result, we define, for $0 < q$ with $q \neq 1$, two functions by

$$\ln_q(s) := \frac{s^{1-q} - 1}{1-q}, \ s \geq 0, \quad \exp_q(t) := [1 + (1-q)t]_+^{1/(1-q)}, \ t \in \mathbf{R},$$

where $[t]_+ := \max\{0, t\}$, and the so-called Tsallis entropy [23] by

$$S_q(\boldsymbol{p}) := \frac{\sum_{i=1}^{n+1}(p_i)^q - 1}{1 - q}.$$

Note that $s = \exp_q(\ln_q(s))$ holds and they respectively recover the usual logarithmic, exponential function and the Boltzmann-Gibbs-Shannon entropy $-\sum_{i=1}^{n+1} p_i \ln p_i$ when $q \to 1$. For $q > 0$, $\ln_q(s)$ is concave on $s > 0$.

Theorem 1 *For the dually flat space $(\mathcal{S}^n, h, \nabla, \nabla^*)$ defined via ± 1-conformal transformation from $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$, the associated potential functions ψ, ψ^*, and dually flat affine coordinate systems $(\theta^1, \cdots, \theta^n)$ and $(\eta_1, \cdots, \eta_n)$ are represented as follows:*

$$\begin{aligned}
\theta^i(\boldsymbol{p}) &= x_i^*(\boldsymbol{p}) - x_{n+1}^*(\boldsymbol{p}), \quad i = 1, \cdots, n \\
\eta_i(\boldsymbol{p}) &= P_i(\boldsymbol{p}), \quad i = 1, \cdots, n \\
\psi(\boldsymbol{\theta}(\boldsymbol{p})) &= -\ln_q(p_{n+1}), \\
\psi^*(\boldsymbol{\eta}(\boldsymbol{p})) &= \frac{1}{\kappa}(\lambda(\boldsymbol{p}) - q) = \frac{1}{1-q}\left(\sum_{i=1}^{n+1}(\eta_i)^{1/q}\right)^q - \frac{1}{1-q},
\end{aligned}$$

where $\kappa = (1 - \alpha^2)/4 = q(1 - q)$ is the scalar curvature of $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ and $\eta_{n+1} := P_{n+1}(\boldsymbol{p}) = 1 - \sum_{i=1}^n P_i(\boldsymbol{p})$. Further, the coordinate systems $(\theta^1, \cdots, \theta^n)$ and $(\eta_1, \cdots, \eta_n)$ are ∇- and ∇^*-affine, respectively.

Proof) As is mentioned above we have only to check that the potential functions ψ, ψ^* and dual affine coordinates θ^i, η_i in the statement satisfy (14) and (15) for the conformal divergence ρ. First, substitute them directly to the right-hand side of (15) and modify it caring for the relation $\eta_{n+1} = 1 - \sum_{i=1}^n \eta_i$, then we see that it coincides with $\rho(\boldsymbol{p}, \boldsymbol{r})$ in (13). Next, since it holds that $\ln_q(p_i) = x_i^*(\boldsymbol{p}) - 1/(1-q)$, we can alternatively represent

$$\theta^i(\boldsymbol{p}) = \ln_q(p_i) - \ln_q(p_{n+1}) = \ln_q(p_i) + \psi(\boldsymbol{\theta}(\boldsymbol{p})), \quad i = 1, \cdots, n.$$

Hence, for $\theta^{n+1} \equiv 0$ it holds

$$1 = \sum_{i=1}^{n+1} p_i = \sum_{i=1}^{n+1} \exp_q(\theta^i - \psi).$$

Differentiating the both sides by θ^j, $j = 1, \cdots, n$, we have

$$0 = \sum_{i=1}^{n+1}\left(\delta_{ij} - \frac{\partial \psi}{\partial \theta^j}\right)(p_i)^q = (p_j)^q - \frac{\partial \psi}{\partial \theta^j}\sum_{i=1}^{n+1}(p_i)^q, \quad j = 1, \cdots, n.$$

Thus, the left equation of (14) holds. Finally, note that the conformal factor is represented by

$$\lambda(\boldsymbol{p}) = \frac{1}{Z_q(\boldsymbol{p})} = \frac{q}{\sum_{i=1}^{n+1}(p_i)^q} = \frac{q}{(\exp_q(S_q(\boldsymbol{p})))^{1-q}}. \tag{16}$$

Using the formula [24]:

$$\exp_q(S_q(\boldsymbol{p})) = \exp_{\frac{1}{q}}\left(S_{\frac{1}{q}}(\boldsymbol{P})\right),$$

we see that

$$\lambda(\boldsymbol{p}) = q \left(\exp_{\frac{1}{q}} \left(S_{\frac{1}{q}}(\boldsymbol{P}) \right) \right)^{q-1} = q \left(\sum_{i=1}^{n+1} (P_i)^{\frac{1}{q}} \right)^q.$$

Hence, the second equality in the expression of ψ^* holds. The right equation of (14) follows if you again recall $\eta_{n+1} = 1 - \sum_{i=1}^{n} \eta_i$.   Q.E.D.

Corollary 1 *The escort probabilities $P_i, i = 1, \cdots, n$ are canonical affine coordinates of the flat affine connection ∇^* on $\mathcal{S}^n$.*

Remark 1: Since the conformal factor λ in (16) can be alternatively represented by

$$\lambda(\boldsymbol{p}) = \frac{q}{(\exp_q(S_q(\boldsymbol{p})))^{1-q}} = \kappa \ln_q \left(\frac{1}{\exp_q(S_q(\boldsymbol{p}))} \right) + q,$$

we have another expression of ψ^*, i.e,

$$\psi^* = \ln_q \left(\frac{1}{\exp_q(S_q(\boldsymbol{p}))} \right).$$

Thus, the potentials and dual coordinates given in the proposition recover the standard ones [3, 4] when $q \to 1$, i.e,

$$\psi \to -\ln p_{n+1}, \quad \psi^* \to \sum_{i=1}^{n+1} p_i \log p_i \quad \theta^i \to \log(p_i/p_{n+1}), \quad \eta_i \to p_i, \quad i = 1, \cdots, n.$$

Note that $-\psi^*$ coincides with the entropy studied in [25, 26, 27] and referred to as the *normalized Tsallis entropy*. The conformal (or scaling) factor λ often appears in the study of the q-analysis.

Remark 2: Similarly to the above conformal transformation of $(\mathcal{S}^n, g, \nabla^{(\alpha)})$, we can define another one for $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$ with a conformal factor

$$\lambda'(\boldsymbol{p}) := \frac{1}{\sum_{i=1}^{n+1} L^{(-\alpha)}(p_i)},$$

and construct another dually flat structure $(h' = \lambda'g, \nabla', \nabla'^*)$. Hence, the following relations among them hold (See Figure 1).

	$(\mathcal{S}^n, h', \nabla')$	$\overset{\text{dual}}{\longleftrightarrow}$	$(\mathcal{S}^n, h', \nabla'^*)$	
1-conformally equivalent	$\updownarrow$		$\updownarrow$	-1-conformally equivalent
	$(\mathcal{S}^n, g, \nabla^{(\alpha)})$	$\overset{\text{dual}}{\longleftrightarrow}$	$(\mathcal{S}^n, g, \nabla^{(-\alpha)})$	
-1-conformally equivalent	$\updownarrow$		$\updownarrow$	1-conformally equivalent
	$(\mathcal{S}^n, h, \nabla^*)$	$\overset{\text{dual}}{\longleftrightarrow}$	$(\mathcal{S}^n, h, \nabla)$	

Figure 1. Relations among geometries

Remark 3: Because of the projective equivalence (11), a submanifold in $\mathcal{S}^n$ is $\nabla^{(\alpha)}$-autoparallel if and only if it is ∇^*-autoparallel. In particular, the set of distributions constrained with the normalized q-expectations (escort averages) [2] is a simultaneously $\nabla^{(\alpha)}$- and ∇^*-autoparallel submanifold in $\mathcal{S}^n$.

4. Applications to construction of alpha-Voronoi diagrams and alpha-centroids

For given m points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ we define α-*Voronoi regions* on $\mathcal{S}^n$ using the α-divergence as follows:

$$\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_k) < D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_l)\}, \quad k = 1, \cdots, m.$$

An α-*Voronoi diagram* on $\mathcal{S}^n$ is a collection of the α-Voronoi regions and their boundaries. Note that $D^{(\alpha)}$ approaches the Kullback-Leibler divergence if $\alpha \to -1$, and $D^{(0)}$ is called the Hellinger distance. If we use the *Rényi divergence of order $\alpha \neq 1$* [28] defined by

$$D_\alpha(\boldsymbol{p}, \boldsymbol{r}) := \frac{1}{\alpha - 1} \ln \sum_{i=1}^{n+1} (p_i)^\alpha (r_i)^{1-\alpha},$$

instead of the α-divergence, $\mathrm{Vor}^{(1-2\alpha)}(\boldsymbol{p}_k)$ gives the corresponding Voronoi region because of their one-to-one functional relationship.

The standard algorithm using projection of a polyhedron [29, 6] commonly works well to construct Voronoi diagrams for the Euclidean distance [6], the Kullback-Leibler [11] and Bregman divergences [12], respectively. The algorithm is applicable if a distance function is represented by the remainder of the first order Taylor expansion of a convex potential function in a suitable coordinate system. Geometrically speaking, this is satisfied if i) the divergence is a canonical one for a certain dually flat structure and ii) its affine coordinate system is chosen to realize the corresponding Voronoi diagrams. In this coordinate system with one extra complementary coordinate the polyhedron is expressed as the upper envelop of m hyperplanes tangent to the potential function.

A problem for the case of the α-Voronoi diagram is that the α-divergence on $\mathcal{S}^n$ *cannot* be represented as a remainder of any convex potentials. The following theorem, however, claims that the problem is resolved by conformally transforming the α-geometry to the dually flat structure (h, ∇, ∇^*) and using the conformal divergence ρ and escort probabilities as a coordinate system.

Here, we denote the point on $\mathcal{E}^n$ by $\boldsymbol{P} = (P_1, \cdots, P_n)$ because $P_{n+1} = 1 - \sum_{i=1}^n P_i$.

Theorem 2 *i) The bisector of $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$ defined by $\{\boldsymbol{p} | D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_k) = D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_l)\}$ is a simultaneously $\nabla^{(\alpha)}$- and ∇^*-autoparallel hypersurface on $\mathcal{S}^n$.*

  ii) Let $\mathcal{H}_k, k = 1, \cdots, m$ be the hyperplane in $\mathcal{E}^n \times \mathbf{R}$ which is respectively tangent at $(\boldsymbol{P}_k, \psi^*(\boldsymbol{P}_k))$ to the hypersurface $\{(\boldsymbol{P}, y) | y = \psi^*(\boldsymbol{P})\}$, where $\boldsymbol{P}_k = \boldsymbol{P}(\boldsymbol{p}_k)$. The α-Voronoi diagram can be constructed on $\mathcal{E}^n$ as the projection of the upper envelope of $\mathcal{H}_k$'s along the y-axis.

Proof) i) Consider the $\nabla^{(-\alpha)}$-geodesic $\gamma^{(-\alpha)}$ connecting $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$, and let $\bar{\boldsymbol{p}}$ be the midpoint on $\gamma^{(-\alpha)}$ satisfying $D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_k) = D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_l)$. Denote by $\mathcal{B}$ the $\nabla^{(\alpha)}$-autoparallel hypersurface that is orthogonal to $\gamma^{(-\alpha)}$ and contains $\bar{\boldsymbol{p}}$. Then, for all
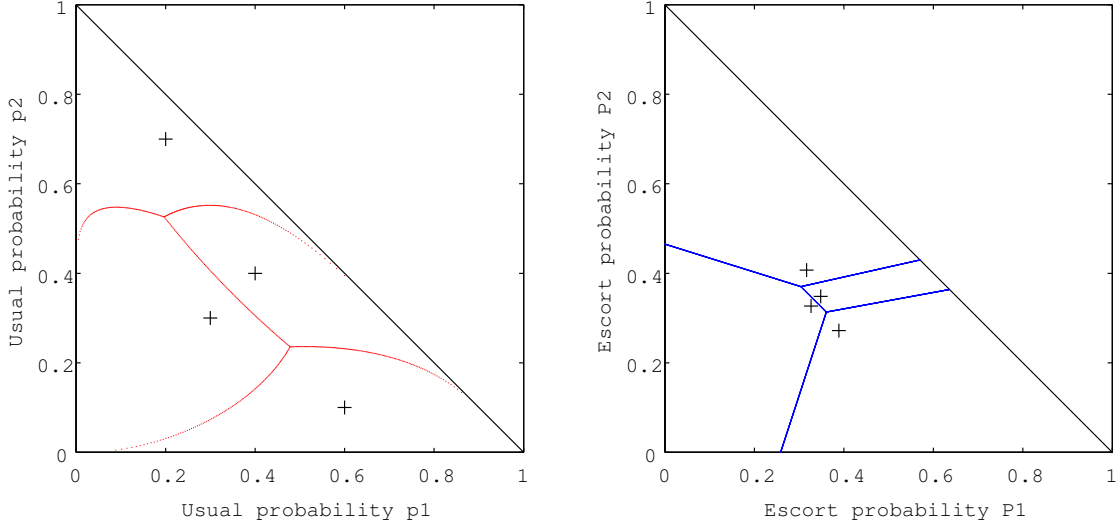
Figure 2. An example of α-Voronoi diagram on $\mathcal{S}^2$ (left) for $\alpha = 0.6$ (or $q = 0.2$) and the corresponding one on $\mathcal{E}^2$ (right).
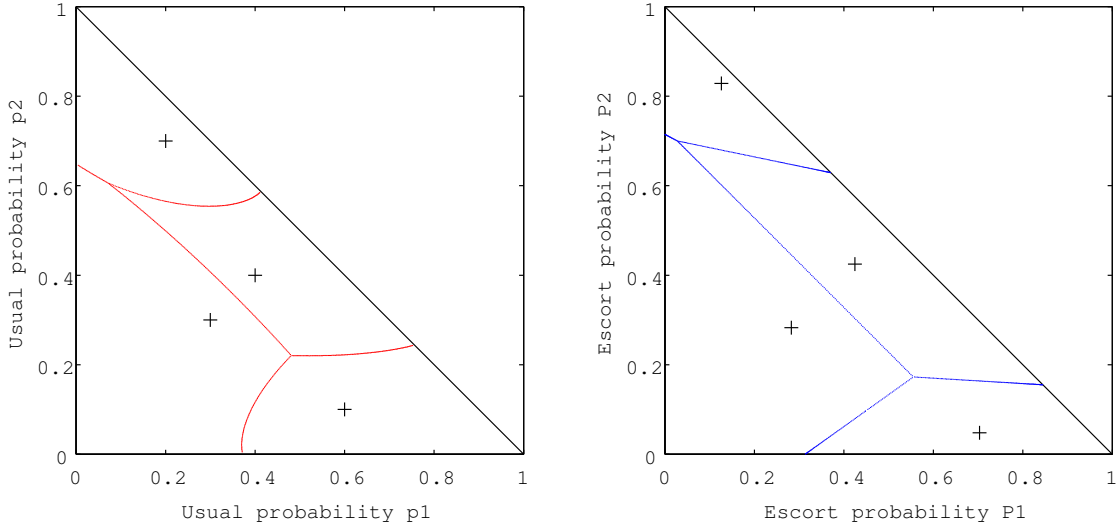

Figure 3. An example of α-Voronoi diagram on $\mathcal{S}^2$ (left) for $\alpha = -2$ (or $q = 1.5$) and the corresponding one on $\mathcal{E}^2$ (right).

$r \in \mathcal{B}$, the modified Pythagorean theorem [20, 7] implies the following equality:

$$D^{(\alpha)}(\boldsymbol{r}, \boldsymbol{p}_k) = D^{(\alpha)}(\boldsymbol{r}, \bar{\boldsymbol{p}}) + D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_k) - \kappa D^{(\alpha)}(\boldsymbol{r}, \bar{\boldsymbol{p}}) D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_k)$$
$$= D^{(\alpha)}(\boldsymbol{r}, \bar{\boldsymbol{p}}) + D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_l) - \kappa D^{(\alpha)}(\boldsymbol{r}, \bar{\boldsymbol{p}}) D^{(\alpha)}(\bar{\boldsymbol{p}}, \boldsymbol{p}_l) = D^{(\alpha)}(\boldsymbol{r}, \boldsymbol{p}_l).$$

Hence, $\mathcal{B}$ is a bisector of $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$. The projective equivalence ensures that $\mathcal{B}$ is also ∇^*-autoparallel.

ii) Recall the equality $D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = D^{(-\alpha)}(\boldsymbol{r}, \boldsymbol{p})$ and the conformal relation (13) between $D^{(-\alpha)}$ and ρ, then we see that $\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k) = \mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k)$ holds on $\mathcal{S}^n$, where

$$\mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | \rho(\boldsymbol{p}_k, \boldsymbol{p}) < \rho(\boldsymbol{p}_l, \boldsymbol{p})\}.$$

Theorem 1, relations (14) and (15) imply that $\rho(\boldsymbol{p}_k, \boldsymbol{p})$ is represented with the coordinates (P_i) by

$$\rho(\boldsymbol{p}_k, \boldsymbol{p}) = \psi^*(\boldsymbol{P}) - \left(\psi^*(\boldsymbol{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\boldsymbol{P}_k)(P_i(\boldsymbol{p}) - P_i(\boldsymbol{p}_k)) \right),$$

where $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{p})$. Note that a point $(\boldsymbol{P}, y_k(\boldsymbol{P}))$ in $\mathcal{H}_k$ is expressed by

$$y_k(\boldsymbol{P}) := \psi^*(\boldsymbol{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\boldsymbol{P}_k)(P_i(\boldsymbol{p}) - P_i(\boldsymbol{p}_k)).$$

Hence, we have $\rho(\boldsymbol{p}_k, \boldsymbol{p}) = \psi^*(\boldsymbol{P}) - y_k(\boldsymbol{P})$. We see, for example, that the bisector on $\mathcal{E}^n$ for $\boldsymbol{p}_k$ and $\boldsymbol{p}_l$ is represented as a projection of $\mathcal{H}_k \cap \mathcal{H}_l$. Thus, the statement follows. Q.E.D.

The figure 2 and 3 show examples of α-Voronoi diagrams on the simplex of dimension 2. In these cases, the bisectors are simultaneously $\nabla^{(\alpha)}$- and ∇^*-geodesics.

Remark 4: In [30] Voronoi diagrams for broader class of divergences (contrast functions) that are not necessarily associated with any convex potentials are studied from more general affine differential geometric points of views. The construction algorithm is also given there, which is applicable if the corresponding affine immersion is explicitly obtained.

On the other hand, the α-divergence defined not only on $\mathcal{S}^n$ but on the positive orthant $\mathbf{R}_+^{n+1}$ can be represented as a remainder of the potential Ψ in (7) [3, 4, 8]. Hence, the α-geometry on $\mathbf{R}_+^{n+1}$ is dually flat. Using this property, α-Voronoi diagrams on $\mathbf{R}_+^{n+1}$ is discussed in [31].

While both of the above methods require computation of the polyhedrons in the space of dimension $n + 2$, the new one proposed in this paper does in the space of dimension $n + 1$. Since the optimal computational time of polyhedrons depends on the dimension d by $O(m \log m + m^{\lfloor d/2 \rfloor})$ [32], the new one where $d = n+1$ is slightly better when n is even.

The next proposition is a simple and relevant application of escort probabilities. Define the α-*centroid* $\boldsymbol{c}^{(\alpha)}$ for given m points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ by the minimizer of the following problem:

$$\min_{\boldsymbol{p} \in \mathcal{S}^n} \sum_{k=1}^m D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}).$$

Proposition 1 *The α-centroid $\boldsymbol{c}^{(\alpha)}$ for given m points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ is represented in escort probabilities by the weighted average of conformal factors $\lambda(\boldsymbol{p}_k) = 1/Z_q(\boldsymbol{p}_k)$, i.e.,*

$$P_i(\boldsymbol{c}^{(\alpha)}) = \frac{1}{\sum_{k=1}^m Z_q(\boldsymbol{p}_k)} \sum_{k=1}^m Z_q(\boldsymbol{p}_k) P_i(\boldsymbol{p}_k), \quad i = 1, \cdots, n+1.$$

Proof) Let $\theta^i = \theta^i(\boldsymbol{p})$. Using (13), (15) and the relation $D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = D^{(-\alpha)}(\boldsymbol{r}, \boldsymbol{p})$, we have

$$\sum_{k=1}^m D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}) = \sum_{k=1}^m Z_q(\boldsymbol{p}_k)\rho(\boldsymbol{p}, \boldsymbol{p}_k) = \sum_{k=1}^m Z_q(\boldsymbol{p}_k)\{\psi(\boldsymbol{\theta}) + \psi^*(\boldsymbol{\eta}(\boldsymbol{p}_k)) - \sum_{i=1}^n \theta^i \eta_i(\boldsymbol{p}_k)\}.$$

Then the optimality condition is

$$\frac{\partial}{\partial \theta^i} \sum_{k=1}^{m} D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}) = \sum_{k=1}^{m} Z_q(\boldsymbol{p}_k)(\eta_i - \eta_i(\boldsymbol{p}_k)) = 0, \quad i = 1, \cdots, n,$$

where $\eta_i = \eta_i(\boldsymbol{p})$. Thus, the statement follows from Theorem 1 for $i = 1, \cdots, n$. For $i = n + 1$ it follows from the fact that the sum of the weights is equal to one.   Q.E.D.

5. Concluding remarks

We have considered ± 1-conformal transformations of the α-geometry and obtained dually flat structure $(\mathcal{S}^n, h, \nabla, \nabla^*)$. Further the potential functions and dually flat coordinate systems associated with the structure have been derived. We see that the escort probability naturally appears to play an important role.

From a viewpoint of contrast functions, the geometric structure compatible to the Kullback-Leibler divergence is $(\mathcal{S}^n, g, \nabla^{(1)}, \nabla^{(-1)})$, where g is the Fisher information and $\nabla^{(\pm 1)}$ are respectively the *e-connection* and the *m-connection*. Similarly, the α-divergence (or the Tsallis relative entropy), and the conformal divergence ρ in this note correspond to $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ and $(\mathcal{S}^n, h, \nabla, \nabla^*)$, respectively. They are summarized in Figure 4.

KL divergence	α-divergence	conformal divergence
$(\mathcal{S}^n, g, \nabla^{(1)}, \nabla^{(-1)})$ $\longleftrightarrow$	$(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ $\longleftrightarrow$	$(\mathcal{S}^n, h, \nabla, \nabla^*)$, $(\mathcal{S}^n, h', \nabla', \nabla'^*)$
dually flat	constant curvature κ	dually flat

Figure 4. transformations of dualistic structures

The physical meaning or essence underlying these transformations would be interesting and significant, but is left unclear. (See recent publications [33, 34] for such research directions.)

Finally, we have shown a direct application of the conformal flattening to computation of α-Voronoi diagrams and α-centroids. Escort probabilities are found to work as a suitable coordinate system for the purpose.

Acknowledgments

Appendix A: Statistical manifold and α-conformally equivalence

For details of this appendix see [17, 19, 20, 21, 22]. For a torsion-free affine connection ∇ and a pseudo Riemannian metric g on a manifold $\mathcal{M}$, the triple $(\mathcal{M}, g, \nabla)$ is called a *statistical manifold* if it admits another torsion-free connection ∇^* satisfying

$$X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \tag{A.1}$$

for arbitrary X, Y and Z in $\mathcal{X}(\mathcal{M})$, where $\mathcal{X}(\mathcal{M})$ is the set of all tangent vector fields on $\mathcal{M}$. It is known that $(\mathcal{M}, g, \nabla)$ is a statistical manifold if and only if ∇g is symmetric, i.e., $(\nabla_X g)(Y, Z)$ is symmetric with respect to X, Y and Z. We call ∇ and ∇^* *duals* of each other with respect to g, and $(\mathcal{M}, g, \nabla^*)$ is said the dual statistical manifold of $(\mathcal{M}, g, \nabla)$. The triple of a Riemannian metric and a pair of dual connections (g, ∇, ∇^*) satisfying (A.1) is called a *dualistic structure* on $\mathcal{M}$.

For $\alpha \in \mathbf{R}$, statistical manifolds $(\mathcal{M}, g, \nabla)$ and $(\mathcal{M}, g', \nabla')$ are said to be α-*conformally equivalent* if there exists a positive function ϕ on $\mathcal{M}$ such that

$$
\begin{aligned}
g'(X, Y) \quad &= \phi g(X, Y), \\
g(\nabla'_X Y, Z) &= g(\nabla_X Y, Z) - \frac{1+\alpha}{2} d(\ln \phi)(Z) g(X, Y) \\
&\quad + \frac{1-\alpha}{2} \{d(\ln \phi)(X) g(Y, Z) + d(\ln \phi)(Y) g(X, Z)\}.
\end{aligned}
$$

Statistical manifolds $(\mathcal{M}, g, \nabla)$ and $(\mathcal{M}, g', \nabla')$ are α-conformally equivalent if and only if $(\mathcal{M}, g, \nabla^*)$ and $(\mathcal{M}, g, \nabla'^*)$ are $-\alpha$-conformally equivalent.

A statistical manifold $(\mathcal{M}, g, \nabla)$ is called α-*conformally flat* if it is locally α-conformally equivalent to a flat statistical manifold. Note that -1-conformal equivalence implies projective equivalence. A statistical manifold of dimension greater than three has constant curvature if and only if it is ± 1-conformally flat.

We call a function ρ on $\mathcal{M} \times \mathcal{M}$ a *contrast function* [35] *inducing* (g, ∇, ∇^*) if it satisfies

$$
\begin{aligned}
\rho(p, p) \quad &= 0, \quad p \in \mathcal{M}, & \text{(A.2)} \\
\rho[X||] \quad &= \rho[|Y] = 0, & \text{(A.3)} \\
g(X, Y) \quad &= -\rho[X|Y], & \text{(A.4)} \\
g(\nabla_X Y, Z) &= -\rho[XY|Z], \quad g(Y, \nabla^*_X Z) = -\rho[Y|XZ], & \text{(A.5)}
\end{aligned}
$$

where

$$
\rho[X_1 \cdots X_k | Y_1 \cdots Y_l](p) := (X_1)_p \cdots (X_k)_p (Y_1)_q \cdots (Y_l)_q \rho(p, q)|_{p=q}
$$

for arbitrary $p, q \in \mathcal{M}$ and $X_i, Y_j \in \mathcal{X}(\mathcal{M})$. If $(\mathcal{M}, g, \nabla)$ and $(\mathcal{M}, g', \nabla')$ are 1-conformally equivalent, a contrast function ρ' inducing (g', ∇', ∇'^*) is represented by ρ inducing (g, ∇, ∇^*), as

$$
\rho'(p, q) = \phi(q) \rho(p, q).
$$

Appendix B: The proof for the fact that $(\mathcal{S}^n, h, \nabla^*)$ is a statistical manifold

We show that $\nabla^* h$ is symmetric. By the definition of -1-conformal flatness we have

$$
\begin{aligned}
(\nabla^*_X h)(Y, Z) &= X h(Y, Z) - h(\nabla^*_X Y, Z) - h(Y, \nabla^*_X Y) \\
&= d\lambda(X) g(Y, Z) + \lambda X g(Y, Z) \\
&\quad - \lambda \{g(\nabla^{(\alpha)}_X Y, Z) + d(\ln \lambda)(Y) g(X, Z) + d(\ln \lambda)(X) g(Y, Z)\} \\
&\quad - \lambda \{g(Y, \nabla^{(\alpha)}_X Z) + d(\ln \lambda)(Z) g(X, Y) + d(\ln \lambda)(X) g(Z, Y)\}.
\end{aligned}
$$

Substitute the equality $\lambda d(\ln \lambda) = d\lambda$ into the right-hand side, then it is transformed to

$$\lambda \{ Xg(Y,Z) - g(\nabla_X^{(\alpha)} Y, Z) - g(Y, \nabla_X^{(\alpha)} Z)$$
$$\qquad - d(\ln \lambda)(X) g(Y,Z) - d(\ln \lambda)(Y) g(X,Z) - d(\ln \lambda)(Z) g(X,Y) \}$$
$$= \lambda (\nabla_X^{(\alpha)} g)(Y,Z) - \lambda \{ d(\ln \lambda)(X) g(Y,Z) + d(\ln \lambda)(Y) g(X,Z) + d(\ln \lambda)(Z) g(X,Y) \}.$$

Thus, $\nabla^* h$ is symmetric because $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ is a statistical manifold, i.e., $\nabla^{(\alpha)} g$ is symmetric. Since $\nabla^{(\alpha)}$ is torsion-free, so is ∇^* by the definition of -1-conformally flatness.

References

[1] Beck C and Schlögl F 1993 *Thermodynamics of Chaotic Systems* (Cambridge: Cambridge University Press)

[2] Tsallis C 2009 *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Berlin/Heidelberg: Springer)

[3] Amari S-I 1985 *Differential-Geometrical Methods in Statistics*, (Berlin Springer-Verlag)

[4] Amari S-I and Nagaoka H 2000 *Methods of Information Geometry*, (Rhode Island: AMS&Oxford).

[5] Nomizu K and Sasaki T 1993 *Affine Differential Geometry* (Cambridge: Cambridge University Press)

[6] Edelsbrunner H 1987 *Algorithms in Combinatorial Geometry* (Springer-Verlag)

[7] Ohara A 2007 Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization *Phys. Lett. A* **370** 184–93

[8] Ohara A 2009 Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation *Euro. Phys. J. B* **70** 15–28

[9] Okamoto I, Amari S and Takeuchi K 1991 Asymptotic theory of sequential estimation procedures for curved exponential families *Annals of Statistics* **19** 961–81

[10] Onishi K and Takayama N 1996 Construction of Voronoi diagram on the Upper Half-Plane *IEICE Trans. on Fundamentals* **E79-A** 533–9

[11] Onishi K and Imai H 1997 Voronoi diagram in statistical parametric space by Kullback-Leibler divergence *Proc. 13th ACM-SIAM Symposium on Computational Geometry* 463–5

[12] Nielsen F, Boissonnat J-D and Nock N 2007 On Bregman Voronoi diagram *Proc. 18th ACM-SIAM Symposium on Discrete Algorithms* 746–55

[13] Imai H and Inaba M 2000 Divergence-Based Geometric Clustering and Its Underlying Discrete Proximity Structures *IEICE Trans. Inf. & Syst.* **E83-D** 27–35

[14] Lebanon G 2005 Metric Learning for Text Documents *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **28**, 497–508

[15] Amari S-I 2007 Integration of stochastic models by minimizing α-divergence Neural Comput., vol. 19, no. 10, 2780–2796

[16] Čencov N N, 1982 *Statistical Decision Rules and Optimal Inference*, AMS, Rhode Island (Originally published in Russian Nauka, Moscow (1972)).

[17] Lauritzen S L 1987 Statistical manifolds S.Amari *et al* eds. *Differential geometry in statistical inference* 163–216 (Hayward: Institute of Mathematical Statistics)

[18] Matsuzoe H, Takeuchi J and Amari S-I 2006 Equiaffine structures on statistical manifolds and Bayesian statistics *Differential Geom. Appl.* **24** 567–78

[19] Kurose T 1990 Dual Connections and Affine Geometry *Math. Z.* **203** 115–21

[20] Kurose T 1994 On the Divergence of 1-conformally Flat Statistical Manifolds, *Tôhoku Math. J.* **46** 427–33

[21] Kurose T 2002 Conformal-projective geometry of statistical manifolds. *Interdiscip. Inform. Sci.* **8** 89–100

[22] Matsuzoe H 1999 Geometry of contrast functions and conformal geometry *Hiroshima Math. J.* **29** 175–91

[23] Tsallis C 1988 Possible generalization of Boltzmann-Gibbs statistics *J. Stat. Phys.* **52**, 479–87

[24] Suyari H and Wada T 2008 Multiplicative duality, q-triplet and (μ, ν, q)-relation derived from the one-to-one correspondence between the (μ, ν)-multinomial coefficient and Tsallis entropy S_q *Physica A 387* 71–83

[25] Landsberg P T and Vedral V 1998 Distributions and channel capacities in generalized statistical mechanics *Phys. Lett. A* **247** 211–7

[26] Rajagopal A K and Abe S 1999 Implications of Form Invariance to the Structure of Nonextensive Entropies *Phys. Rev. Lett.* **83** 1711–4

[27] Wada T and Scarfone A M 2005 Connections between Tsallis' formalisms employing the standard linear average energy and ones employing the normalized q-average energy *Phys. Lett. A* **335** 351–62

[28] Rényi A 1961 On measures of entropy and information *Proc. 4th Berkeley Symp. Math. Statist. Probability* **1** 547–61 (Univ. Calif. Press: Berkeley)

[29] Edelsbrunner H and Seidel R 1986 *Discrete Comput. Geom.* **1** 25–44

[30] Matsuzoe H 2009 Computational Geometry from the Viewpoint of Affine Differential Geometry *F.Nielsen eds. Emerging Trends in Visual Computing* 103–23 (Berlin/Heidelberg: Springer)

[31] Nielsen, F and Nock, R 2009 The dual Voronoi diagrams with respect to representational Bregman divergences. International Symposium on Voronoi Diagrams (ISVD), DTU Lyngby, Denmark, June 2009. IEEE Press, New York

[32] Chazelle B, 1993 An optimal convex hull algorithm in any fixed dimension, *Discrete and Computational Geometry* **10** 377–409

[33] Ojima I 2010 Meaning of Non-Extensive Entropies in Micro-Macro Duality *J. Phys.: Conf. Ser.* **201** 012017

[34] Tanaka M 2010 Meaning of an escort distribution and τ-transformation *J. Phys.: Conf. Ser.* **201** 012007

[35] Eguchi S 1992 Geometry of minimum contrast *Hiroshima Math. J.* **23** 631–47

[36] Ohara A. Matsuzoe H and Amari S 2010 A dually flat structure on the space of escort distributions *J. Phys.: Conf. Ser.* **201** 012012